

УТВЕРЖДЕНА

Решением Ученого совета
АНО ВО «Центральный университет»
«07» марта 2024 г.
Протокол №1

**Рабочая программа дисциплины (модуля)
«Reinforcement Learning (Обучение с подкреплением)»**

Направление подготовки: 02.03.01 Математика и компьютерные науки

Направленность (профиль) подготовки: Искусственный интеллект

Квалификация (степень) выпускника: бакалавр

Форма обучения: очная

Срок освоения программы: 4 года

Год набора: 2024

**Москва
2024**

Содержание

1. Краткая характеристика дисциплины (модуля)	3
2. Перечень планируемых результатов обучения	4
3. Тематический план	6
4. Содержание дисциплины (модуля)	6
5. Учебно-методическое обеспечение	7
6. Материально-техническое обеспечение	7
7. Методические и оценочные материалы	9

1. Краткая характеристика дисциплины (модуля)

Рабочая программа дисциплины (модуля) «Reinforcement Learning (Обучение с подкреплением)» составлена в соответствии с федеральным государственным образовательным стандартом высшего образования – бакалавриат по специальности 02.03.01 Математика и компьютерные науки, профиль Искусственный интеллект, утвержденный приказом Министерства науки и высшего образования Российской Федерации № 807 от 23.08.2017 года.

Изучение дисциплины (модуля) «Reinforcement Learning (Обучение с подкреплением)» позволяет создавать адаптивные системы и решения в различных прикладных задачах, включая робототехнику, игры и управление; развивает навыки математического моделирования и программной реализации сложных алгоритмов, востребованных в современной науке и индустрии.

Место дисциплины (модуля) в структуре образовательной программы

Настоящая дисциплина (модуль) включена в учебный план по программе подготовки бакалавриата по направлению 02.03.01 Математика и компьютерные науки, профиль Искусственный интеллект и входит в вариативную часть Блока 1, формируемую участниками образовательных отношений как дисциплина по выбору.

Дисциплина (модуль) изучается на 3 или 4 курсе в 5, 6, 7 или 8 семестре на выбор, доступна для прохождения при условии успешного завершения дисциплин «Deep Learning (Глубокое обучение)».

Цель изучения дисциплины (модуля): освоение основных принципов и алгоритмов обучения с подкреплением для разработки интеллектуальных агентов, способных принимать оптимальные решения в динамических средах.

Задачи изучения дисциплины (модуля):

- формирование знаний основных марковских процессов принятия решений (MDP);
- формирование знаний алгоритмов обучения с подкреплением (Q-Learning, Policy Gradients, DQN);
- формирование знаний применение в робототехнике, играх и автоматизации;
- формирование умения реализации RL-алгоритмов на Python (PyTorch/TensorFlow, OpenAI Gym);
- формирование умения анализ и интерпретация результатов обучения агентов;
- формирование навыков разработки RL-агентов для конкретных задач;
- формирование навыков оптимизации гиперпараметров и развертывание решений.

2. Перечень планируемых результатов обучения

Компетенции, формируемые в результате освоения дисциплины (модуля) при проведении учебных занятий в форме контактной работы обучающихся с педагогическими работниками Университета и в форме самостоятельной работы обучающихся:

Компетенция	Содержание компетенции	Индикатор компетенции	Перечень планируемых результатов обучения по дисциплине (модулю)
ОПК-1.	Способен находить, формулировать и решать актуальные и значимые проблемы прикладной и компьютерной математики	ОПК-1.1.	Знает основные методы и подходы к решению задач прикладной и компьютерной математики, включая алгоритмы, математическое моделирование и теорию оптимизации, а также современные инструменты и технологии, используемые в этой области
		ОПК-1.2.	Умеет анализировать и формулировать математические задачи, применять соответствующие методы и алгоритмы для их решения, а также интерпретировать и представлять результаты в понятной и доступной форме
		ОПК-1.3.	Имеет практический опыт работы над проектами или исследованиями в области прикладной и компьютерной математики, включая участие в конкурсах, олимпиадах или научных публикациях, где были решены актуальные и значимые задачи
ОПК-4.	Способен находить, анализировать, реализовывать программно и использовать на практике математические алгоритмы, в том числе с применением современных вычислительных систем	ОПК-4.1.	Знает базовые основы современного математического аппарата, связанного с проектированием, разработкой, реализацией и оценкой качества программных продуктов и программных комплексов в различных областях человеческой деятельности.
		ОПК-4.2.	Умеет использовать этот математический аппарат в профессиональной деятельности.
		ОПК-4.3.	Имеет практический опыт применения современного математического аппарата, связанного с проектированием, разработкой, реализацией и оценкой качества программных продуктов и программных комплексов в различных

			областях человеческой деятельности.
ОПК-6.	Способен разрабатывать алгоритмы и компьютерные программы, пригодные для практического применения	ОПК-6.1.	Знает алгоритмы разработки, компьютерные программы, а также алгоритмы вычислительной математики в области искусственного интеллекта
		ОПК-6.2.	Умеет разрабатывать математические программные продукты и комплексы с использованием современных технологий программирования в области искусственного интеллекта
		ОПК-6.3.	Имеет практический опыт разработки интеллектуальных информационных систем для визуализации результатов исследований в области искусственного интеллекта
ПК-3.	Способен применять методы математического и алгоритмического моделирования для решения как теоретических, так и практических задач в рамках профессиональной деятельности	ПК-3.1.	Знает основные методы математического и алгоритмического моделирования, а также их применение для решения теоретических и прикладных задач
		ПК-3.2.	Умеет применять методы математического и алгоритмического моделирования для анализа и решения различных задач в области математики и компьютерных наук
		ПК-3.3.	Имеет опыт использования методов математического и алгоритмического моделирования при решении теоретических и прикладных задач в профессиональной деятельности

3. Тематический план

№п/п	Наименование раздела дисциплины (модуля)	Трудоемкость, академические часы				ТКУ (текущий контроль успеваемости)
		Очная форма				
		Контактная работа		Контроль	Самостоятельная работа	
Лекции	Семинары (практические занятия)					
1	Основы Reinforcement Learning	10	10		42	Домашние задания
2	Классические алгоритмы RL	10	10		42	Домашние задания
3	Продвинутые методы и приложения	10	10		40	Домашние задания
	<i>Зачет с оценкой</i>			6		Проект
	Итого:	30	30	6	124	
	Объем дисциплины (модуля) (в ак. ч.)	190				
	Объем дисциплины (модуля) (в зач. ед.)	5				

4. Содержание дисциплины (модуля)

№п/п	Наименование раздела дисциплины (модуля)	Содержание дисциплины (модуля) по темам
1	Основы Reinforcement Learning	Формализация задачи обучения с подкреплением: агенты, среды, состояния, действия, награды. Модели принятия решений: Марковские процессы принятия решений (MDP). Функции ценности и политика: определение и роль в обучении. Основные элементы цикла взаимодействия агента со средой
2	Классические алгоритмы RL	Метод динамического программирования: итерация по значениям и политикам. Модельные методы обучения: методы оценки и улучшения политики. Модель-независимые методы: Temporal Difference (TD) обучение, Q-обучение. SARSA и другие алгоритмы обучения с подкреплением с онлайн-обновлением
3	Продвинутые методы и приложения	Функциональное аппроксимирование и глубокое обучение в RL (Deep Q-Networks, DQN). Политики на основе градиента и методы актор-критик (Policy Gradient, Actor-Critic). Обучение с подкреплением в непрерывных пространствах и с частичной наблюдаемостью (POMDP). Применение RL в робототехнике, играх, управлении и других прикладных областях

5. Учебно-методическое обеспечение

Университет располагает полным набором лицензионного и свободно распространяемого программного обеспечения, включая продукты отечественного производства.

Каждый студент в течение всего периода обучения получает индивидуальный неограниченный доступ к электронно-библиотечной системе и электронной информационно-образовательной среде университета. Эти системы предоставляют возможность доступа к ресурсам из любой точки, где есть подключение к сети Интернет, как на территории университета, так и за его пределами.

Студентам обеспечен удаленный доступ к современным профессиональным базам данных и информационным справочным системам.

Основная литература:

1. Саттон Р. С., Барто Э. Дж. Обучение с подкреплением: Введение. 2-е изд. / пер. с англ. А. А. Слинкина. – М.: ДМК Пресс, 2020. – 552 с.:

Дополнительная литература:

1. Высоконагруженные приложения. Программирование, масштабирование, поддержка. — СПб.: Питер, 2018. — 640 с.: ил. — (Серия «Бестселлеры O'Reilly»). ISBN 978-5-4461-0512-0.

6. Материально-техническое обеспечение

Университет располагает материально-технической базой, соответствующей действующим противопожарным правилам и нормам и обеспечивающей проведение всех видов дисциплинарной и междисциплинарной подготовки, практической и научно-исследовательской работ обучающихся, предусмотренных учебным планом.

Помещения, которые представляют собой учебные аудитории для проведения занятий лекционного типа, занятий семинарского (практического) типа, групповых и индивидуальных консультаций, текущего контроля и промежуточной аттестации, а также помещения для самостоятельной работы и помещения для хранения и профилактического обслуживания учебного оборудования. Помещения укомплектованы специализированной мебелью и техническими средствами обучения, служащими для представления учебной информации большой аудитории.

Изучение дисциплины (модуля) обеспечивается в учебных аудиториях, оснащенных:

- столами и стульями;
- компьютерной техникой;
- специализированным оборудованием, включая демонстрационное оборудование.

Помещения для самостоятельной работы обучающихся, в том числе приспособленные для использования инвалидами и лицами с ограниченными возможностями здоровья, оснащены компьютерной техникой с возможностью подключения к сети «Интернет» и обеспечением доступа к в электронную информационно-образовательную среду Университета.

Обучающимся предоставляется доступ (в том числе удаленный) к ресурсам информационно-телекоммуникационной сети «Интернет», электронным ресурсам (в том числе электронным библиотечным системам, современным профессиональным базам данных и информационным справочным системам):

№	Наименование портала (издания, курса, документа)	Ссылка
1.	Научная электронная библиотека elibrary.ru библиотека	https://elibrary.ru/defaultx.asp

2.	База данных для IT-специалистов	https://habr.com
3.	База данных ScienceDirect	https://www.sciencedirect.com
4.	Официальный сайт Министерства науки и высшего образования Российской Федерации	https://minobrnauki.gov.ru/
5.	Федеральный портал «Российское образование»	https://www.edu.ru/
6.	Информационная система "Единое окно доступа к образовательным ресурсам"	http://window.edu.ru/
7.	Единая коллекция цифровых образовательных ресурсов	http://school-collection.edu.ru/
8.	Федеральный центр информационно - образовательных ресурсов	http://fcior.edu.ru/

Перечень информационных технологий, используемых при осуществлении образовательного процесса по дисциплине (модулю), в том числе комплект лицензионного программного обеспечения, современные профессиональные базы данных и информационные справочные системы:

Наименование ПО	Производство	Лицензионное / свободно распространяемое
Операционные системы:		
Microsoft Imagine (Windows Client, Server)	зарубежное	лицензионное
Браузеры:		
Яндекс.Браузер	отечественное	свободно распространяемое
Google Chrome	зарубежное	свободно распространяемое
Офисные приложения:		
Microsoft Imagine (Visio, OneNote)	зарубежное	лицензионное
TeXstudio	зарубежное	свободно распространяемое
Adobe Acrobat Reader	зарубежное	свободно распространяемое
Программное обеспечение для планирования и учета времени:		
Toggle app	зарубежное	свободно распространяемое
Системы управления проектами:		
Microsoft Imagine (Project)	зарубежное	лицензионное
Системы управления базами данных:		
Microsoft Imagine (SQL Server)	зарубежное	лицензионное
Системы резервного копирования (backup):		
Acronis Backup Advanced for HyperV	зарубежное	лицензионное
Справочно-правовые системы:		
КонсультантПлюс: справочно-правовая система	отечественное	лицензионное
Средства антивирусной защиты:		
Kaspersky Endpoint Security для бизнеса Стандартный Russian Edition	отечественное	лицензионное
Среды разработки:		
Visual Studio Code	зарубежное	свободно распространяемое
Bash (Unix shell)	зарубежное	свободно распространяемое
Anaconda	зарубежное	свободно распространяемое
Robotic Operating System	зарубежное	свободно распространяемое
CopelliaSim	зарубежное	свободно распространяемое
Google Colaboratory	зарубежное	свободно распространяемое
Пакеты программных средств и библиотек:		
AutoPsy	зарубежное	свободно распространяемое
Interactive Disassembler (IDA)	зарубежное	свободно распространяемое
Системы управления библиографической информацией:		
Zotero	зарубежное	свободно распространяемое

Сервисы и службы:		
Bind	зарубежное	свободно распространяемое
Docker	зарубежное	свободно распространяемое

7. Методические и оценочные материалы

Методические указания для обучающихся по освоению дисциплины (модуля)

В процессе изучения дисциплины (модуля) «Reinforcement Learning (Обучение с подкреплением)» в рамках текущего контроля успеваемости используются такие виды учебной работы, как лекции, практические занятия, домашние задания, проект, а также различные виды самостоятельной работы обучающихся по заданию преподавателя, направленные на развитие навыков профессиональной лексики, закрепление практических профессиональных компетенций, поощрение инициатив.

Лекция – систематическое, последовательное, монологическое изложение преподавателем учебного материала, как правило, теоретического характера.

В процессе лекций рекомендуется вести конспект лекций: кратко и схематично фиксировать основные идеи, выводы и обобщения лекции; выделять важные мысли, ключевые слова и термины. Необходимо отметить вопросы или материалы, которые вызывают затруднения, и попытаться найти ответы в рекомендованной литературе. Если разобраться в материале не удастся, следует сформулировать вопрос и задать его преподавателю на консультации или во время семинарского (практического) занятия.

Участие в семинаре (практическом занятии) – активная работа студента на семинаре, его ответы на вопросы преподавателя и участие в дискуссии.

Для успешного участия в семинаре студентам рекомендуется заранее ознакомиться с темой обсуждения, прочитать необходимые материалы и подготовить вопросы. Важно активно слушать и вовлекаться в дискуссию, высказывая свои мнения и аргументируя их. При ответах на вопросы преподавателя стоит быть уверенным, четким и логичным, опираясь на изученный материал. Также полезно поддерживать диалог с однокурсниками, чтобы обогатить обсуждение и расширить свои знания.

Домашнее задание – набор задач по темам недели.

При работе над домашними заданиями важно внимательно ознакомиться с требованиями и сроками выполнения. Рекомендуется разбивать задания на этапы, чтобы избежать перегрузки и лучше усвоить материал. Использовать различные источники информации, включая учебники и онлайн-ресурсы, для более глубокого понимания темы.

Проект – исследовательская работа по курсу и презентация результатов.

Для успешной подготовки к проекту: четко определите цели и задачи проекта, распределите роли и обязанности между участниками, а также установите сроки выполнения каждой части работы. Регулярно проводите встречи для обсуждения прогресса и решения возникающих вопросов.

Самостоятельная работа – работа студентов, направленная на углубленное изучение отдельных тем и вопросов учебной дисциплины (модуля).

В процессе самостоятельной работы студенты взаимодействуют с рекомендованными материалами при минимальном участии преподавателя. Задачи студента включают работу с конспектами лекций (обработка текста), повторное изучение учебных материалов планов и тезисов ответов, изучение дополнительных тем, выполнение учебно-исследовательских заданий и другое.

Система оценивания результатов обучения по дисциплине (модулю)

Критерии получения уровня и оценивания сформированности компетенций по дисциплине (модулю) «Reinforcement Learning (Обучение с подкреплением)»

Оценивание уровня учебных достижений, обучающихся по дисциплине (модулю), осуществляется в виде текущего контроля успеваемости и промежуточной аттестации.

Промежуточная аттестация по дисциплине (модулю) осуществляется в форме *зачета с оценкой*, при этом проводится оценка компетенций, сформированных по дисциплине.

Для оценивания текущего контроля успеваемости и промежуточной аттестации используется десятибалльная шкала оценивания, которая соотносится с традиционной пятибалльной шкалой следующим образом:

Десятибалльная оценка	Пятибалльная оценка	Оценка за зачет	Общая характеристика результата обучения по дисциплине (модулю)
10	Отлично	Зачтено	Студент полностью владеет знаниями, изложенными в рабочей программе, и глубоко осмысляет дисциплину. Он самостоятельно и логически последовательно отвечает на все вопросы, акцентируя внимание на наиболее важном. Умеет анализировать, сравнивать, классифицировать, обобщать, конкретизировать и систематизировать изученный материал, выделяя ключевые моменты и устанавливая причинно-следственные связи. Четко формулирует ответы, уверенно интерпретирует результаты анализов и других исследований, а также решает сложные задачи. Студент хорошо знаком с методами исследования, необходимыми для практической деятельности, и умеет связывать теоретические аспекты дисциплины (модуля) с практическими задачами.
9	Отлично	Зачтено	
8	Отлично	Зачтено	
7	Хорошо	Зачтено	Студент обладает знаниями предмета почти в полном объеме рабочей программы и самостоятельно, логически последовательно и всесторонне отвечает на все вопросы, акцентируя внимание на наиболее значимых моментах. Он умеет анализировать, сравнивать, классифицировать, обобщать, конкретизировать и систематизировать изученный материал, выделяя его ключевые аспекты и устанавливая причинно-следственные связи. Формулирует свои ответы, уверенно
6	Хорошо	Зачтено	

Десятибалльная оценка	Пятибалльная оценка	Оценка за зачет	Общая характеристика результата обучения по дисциплине (модулю)
			интерпретирует результаты анализов и других исследований, а также решает сложные ситуационные задачи. Студент хорошо знаком с методами исследования, необходимыми для практической деятельности, и умеет связывать теоретические аспекты предмета с практическими задачами.
5	Удовлетворительно	Зачтено	Студент обладает базовыми знаниями по дисциплине (модулю), но испытывает трудности при самостоятельных ответах и использует неточные формулировки. В ходе ответов он допускает ошибки, касающиеся сути вопросов. Студент способен решать только самые простые задачи и владеет лишь минимальным набором методов исследования.
4	Удовлетворительно	Зачтено	
3	Не сдан	Не зачтено	Студент не овладел обязательным минимумом знаний по предмету и не может ответить на вопросы, даже если преподаватель задает дополнительные наводящие вопросы.
2	Не сдан	Не зачтено	
1	Не сдан	Не зачтено	

Дисциплина (модуль) «Reinforcement Learning (Обучение с подкреплением)» оценивается следующим образом:

Активность	Вес	Описание
Домашние задания	50%	Набор задач по темам недели
Проекты	50%	Исследовательская работа по курсу и презентация результатов

Формула расчёта итоговой оценки по дисциплине (модулю) «Reinforcement Learning (Обучение с подкреплением)»: « $0,5 \times$ среднее за домашние задания + $0,5 \times$ среднее за проект».

Текущий контроль успеваемости обучающихся по дисциплине (модулю)

Примерные домашние задания

Домашнее задание «Основы Reinforcement Learning»

1. Определите ключевые компоненты задачи обучения с подкреплением: агент, среда, состояние, действие и награда. Приведите пример из реальной жизни.
2. Объясните, что такое Марковский процесс принятия решений (MDP) и опишите его основные элементы.
3. В чем состоит разница между функцией ценности состояния $V(s)$ и функцией ценности действия $Q(s, a)$?
4. Опишите, что такое политика π и как она связана с функциями ценности.
5. Нарисуйте и объясните цикл взаимодействия агента и среды в задаче обучения с подкреплением.

Домашнее задание «Классические алгоритмы RL»

1. Опишите метод динамического программирования для решения MDP и объясните, как происходит итерация по значениям.
2. Что такое политика итерация? Как она используется для улучшения качества агента?
3. Объясните идею Temporal Difference (TD) обучения и чем оно отличается от метода Монте-Карло.
4. Опишите алгоритм Q-обучения и приведите пример обновления Q-значения.
5. В чем отличие алгоритмов Q-обучения и SARSA? Приведите пример ситуации, когда SARSA предпочтительнее.

Домашнее задание «Продвинутые методы и приложения»

1. Что такое функциональное аппроксимирование в RL и почему оно необходимо при работе с большими пространствами состояний?
2. Опишите архитектуру Deep Q-Network (DQN) и объясните, как она решает проблему обучения с подкреплением.
3. Объясните основные идеи методов градиентного обучения политики (Policy Gradient).
4. Что такое Actor-Critic методы? Опишите роль актёра и критика в этих алгоритмах.
5. Приведите пример применения глубокого обучения в RL на практике
6. Что такое задачи с частичной наблюдаемостью (POMDP) и как они отличаются от классических MDP?
7. Опишите основные подходы к обучению с подкреплением в непрерывных пространствах состояний и действий.
8. Приведите пример применения RL в робототехнике и опишите, какие задачи решаются с помощью RL в этой области.
9. Как RL применяется в игровых средах? Приведите пример успешного применения.
10. Обсудите вызовы и перспективы использования RL в управлении сложными системами.

Примерное описание к проекту

Проект по теме: «Классические алгоритмы обучения с подкреплением»

Цели проекта:

- Изучить и реализовать классические алгоритмы обучения с подкреплением: итерацию по значениям, итерацию по политикам, Q-обучение и SARSA.
- Понять различия между модельными и модель-независимыми методами RL.
- На практике применить алгоритмы к решению задачи управления в дискретном пространстве состояний и действий.
- Проанализировать эффективность и особенности каждого алгоритма.

Задачи проекта:

1. Теоретически описать основные алгоритмы: итерация по значениям, итерация по политикам, Q-обучение и SARSA.
2. Выбрать среду с дискретным пространством состояний и действий (например, GridWorld или FrozenLake из OpenAI Gym).
3. Реализовать алгоритмы итерации по значениям и итерации по политикам для выбранной среды.
4. Реализовать алгоритмы Q-обучения и SARSA с онлайн-обновлением.

5. Провести сравнение алгоритмов по скорости сходимости, качеству полученной политики и устойчивости к параметрам обучения.

6. Подготовить отчет с описанием реализации, результатами экспериментов и анализом.

Этапы выполнения проекта:

1. Подготовительный этап (1 неделя)

- Изучение теории классических алгоритмов RL.
- Выбор и знакомство со средой для экспериментов.

2. Реализация модельных методов (1–2 недели)

- Итерация по значениям и итерация по политикам.
- Тестирование и отладка кода.

3. Реализация модель-независимых методов (1–2 недели)

- Q-обучение и SARSA.
- Настройка параметров обучения.

4. Эксперименты и анализ (1 неделя)

- Сравнение алгоритмов.
- Визуализация результатов.

5. Подготовка отчета и презентации (1 неделя)

- Описание теории, реализации, результатов.
- Подготовка демонстрации кода.

Критерии оценивания:

Критерий	Максимальный балл	Описание
Теоретическая часть	2	Полнота и корректность описания алгоритмов и методов.
Реализация алгоритмов	2	Работоспособность и качество кода, соответствие требованиям.
Экспериментальный анализ	2	Глубина анализа, качество сравнения алгоритмов, интерпретация результатов.
Отчет и презентация	2	Ясность, структура, оформление, качество презентации.
Демонстрация работы кода	2	Успешный показ работы алгоритмов, ответы на вопросы.

Критерии защиты проекта:

- Четкое и последовательное изложение теоретической базы.

- Демонстрация работы реализованных алгоритмов на выбранной среде.
- Обоснование выбора параметров и методов сравнения.
- Ответы на вопросы по теории и реализации.
- Умение анализировать и интерпретировать полученные результаты.

Задания для промежуточной аттестации по дисциплине (модулю)

№ п/п	Задание	Ответ	Компетенция
1.	Какой из следующих элементов НЕ является частью формализации задачи обучения с подкреплением? а) Агент б) Среда в) Функция активации г) Награда	в	ПК-3
2.	Как называется модель принятия решений, в которой вероятность перехода в следующее состояние зависит только от текущего состояния и действия?	Марковский процесс принятия решений/MDP	ОПК-1
3.	Как называется алгоритм динамического программирования, который обновляет оценки значений состояний до сходимости?	Итерация по значениям	ОПК-4
4.	Какой алгоритм обучения с подкреплением использует онлайнное обновление Q-функции на основе текущего действия и награды, не требуя модели среды?	Q-обучение	ОПК-4
5.	Как называется метод, в котором политика обновляется напрямую с использованием градиента функции вознаграждения?	Политика на основе градиента/Policy Gradient	ОПК-4
6.	Как называется тип среды с подкреплением, где агент имеет неполную информацию о состоянии среды?	Частично наблюдаемое Марковское решение/POMDP	ОПК-6
7.	Как называется подход, при котором агент обучается на основе частичной информации о состоянии среды?	POMDP (Partially Observable Markov Decision Process)	ПК-3
8.	Какой метод используется для оценки и улучшения политики в обучении с подкреплением?	Методы оценки политики	ОПК-1
9.	Какой метод обучения с подкреплением используется для работы с непрерывными пространствами состояний?	Методы актор-критик	ОПК-6
10.	Как называется подход, при котором используется глубокое обучение для аппроксимации функций ценности?	Deep Q-Networks (DQN)	ПК-3